Walmart Sales Forecasting

Aakriti, Akshay, Narae, Niharika, Shweta


How can we forecast the unit sales of various Walmart retail goods?

Current Forecasting is Subjective

- **Relying on domain knowledge**
- **Forecast not backed by data**
- **Ineffective inventory management**

Machine Learning can achieve high forecast accuracy

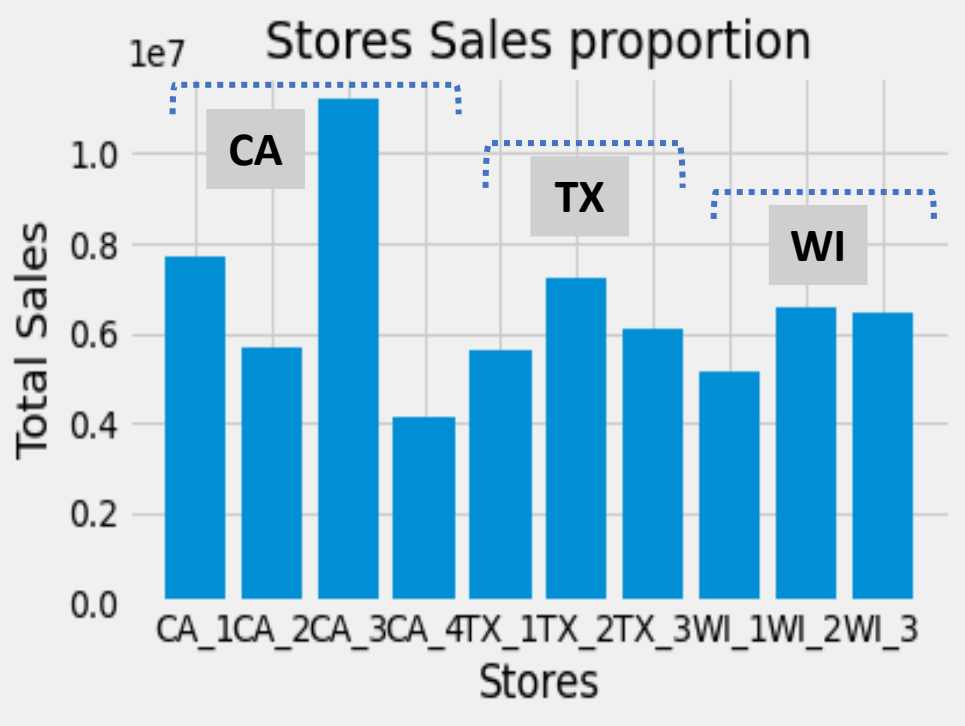
Accurate prediction
Staffing
Inventory

Informed Decision
Based on the data

Increase Revenue
Cut staffing and inventory cost

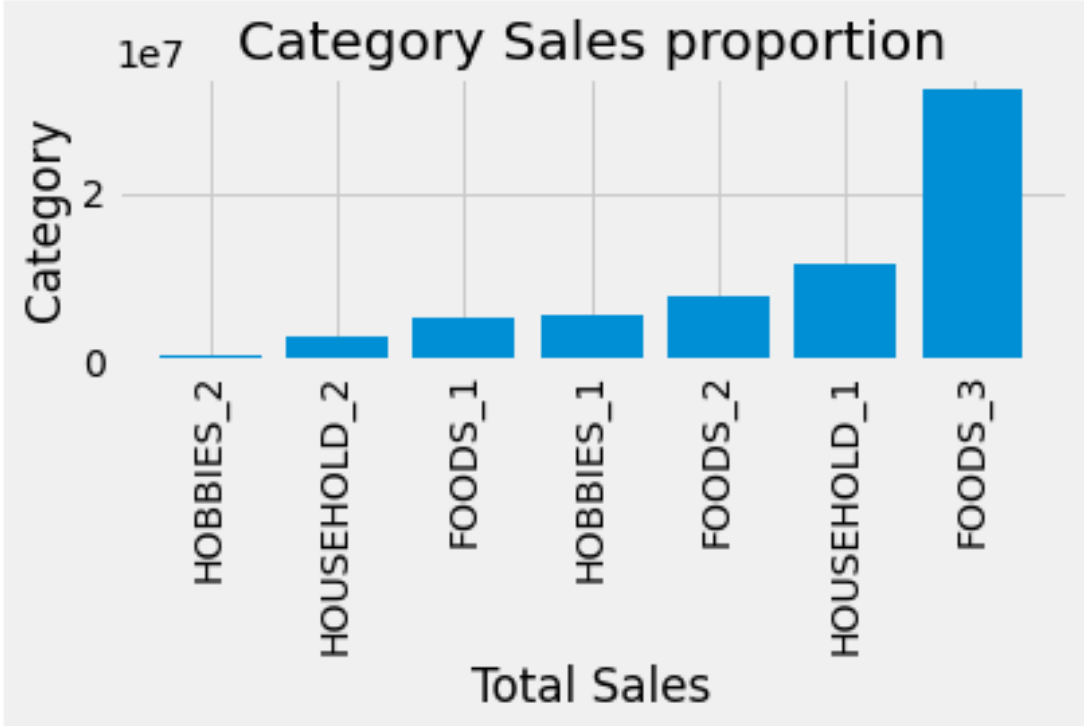
Forecasting on Store and Product Category level

Store Level



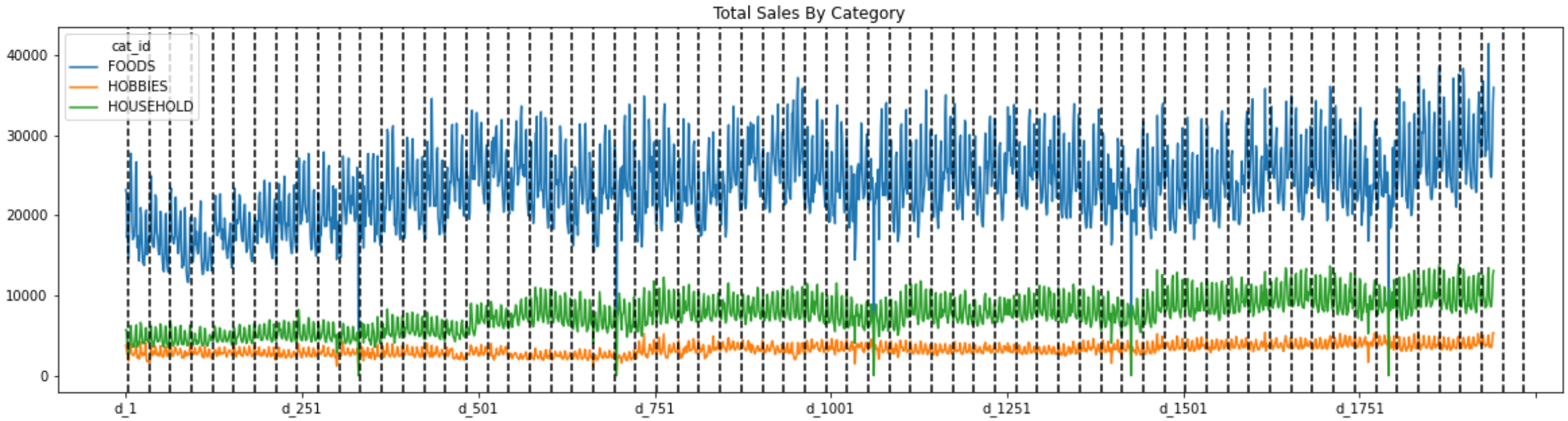
Variance of sales across different stores in each states

Category Level



Variance of sales across different category

Sales pattern across time



Variance
of sales
Across Time

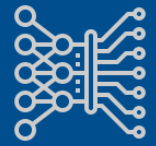
Pattern
of sales
Across Time

Category
differentiate
Sales pattern

Our Solution using Predictive Modeling

Input Data for Forecasting

Historical Product Sales and Time Series Data



Machine Learning Modeling

Using Industry proven techniques for forecasting

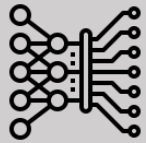
28

Days of Product Unit Sales Predictions

Our Solution using Predictive Modeling

Input Data for Forecasting

Historical Product Sales and Time Series Data



Machine Learning Modeling

Using Industry proven techniques for forecasting

28

Days of Product Unit Sales Predictions

Feature Engineering

Seasonality

Capturing the sales deviations on weekdays, High/Low spending months

Demand Shifts

Creating lag features to identify time series patterns for each product and store

Transformations

Rolling Window statistics to accommodate unusual but expected spikes

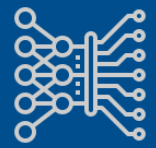
Week over week Fluctuations

Accounting for fluctuation in price week over week for the five-year duration

Our Solution using Predictive Modeling

Input Data for Forecasting

Historical Product Sales and Time Series Data



Machine Learning Modeling

Using Industry proven techniques for forecasting

28

Days of Product Unit Sales Predictions

Modeling Techniques available for Forecasting

BOOSTING

XGBOOST

Extreme Gradient Boosting

Ensemble version of Decision Tree

Scalable, Distributed, Parallel computation

LightGBM

Light Gradient Boosting Machine

Ensemble version of Decision Tree

Utilizes one side sampling and exclusive feature bundling

TIME SERIES

ARIMA

Autoregressive Integrated Moving average

Statistical analysis model

Uses trends in time series data to predict future values

SEQ2SEQ

Sequence to Sequence

Artificial neural network

Uses sequential data for predictions

LSTM

Long Short - Term memory

Recurrent Neural Network

Capable of learning long-term dependencies for sequence predictions

Comparison of the Modeling Techniques

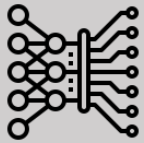
	BOOSTING		TIME SERIES		
	XGBOOST	LightGBM	ARIMA	SEQ2SEQ	LSTM
Measure of Error (RMSE)	1.08	0.61	1.07	1.57	0.82
TIME (mins)	40	45	550	120	40

With best model performance & low computation time, LightGBM is the best model

Our Solution using Predictive Modeling

Input Data for Forecasting

Historical Product Sales and Time Series Data



Machine Learning Modeling

Using Industry proven techniques for forecasting

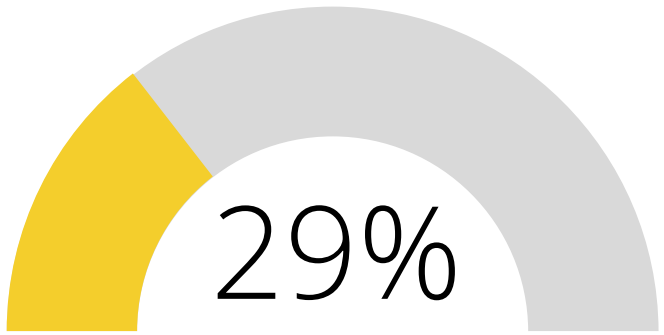
28

Days of Product Unit Sales Predictions

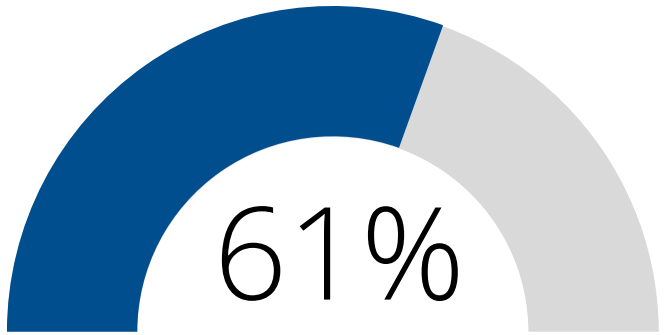
Impact Driven from Our Solution

With this solution, Walmart can save on over & under estimation loss, and increase their revenue by efficient inventory management

If there were no predictions, on average product sales would be overpredicted by



Walmart can earn more revenue with effective product sale prediction by



Our Solution puts us in the top

10%

of the Kaggle Leaderboard

With low error, our solution is able to accurately predict in comparison to other solutions

Thanks for watching